FULLY ENGLISH TRANSLATION OF JP 11-143646

(54) 【Title of the Invention】

Control method for external storage sub-system and external storage sub-system

(57) 【Summary】

【Problem】 Improving throughput of erasing data set in a storage device with a redundant storage configuration which emulates a plurality of logical devices.

【Means of Solution】 In an external storage sub-system where a controller 12 is disposed between a host device 10 and a storage device 13, and the storage device emulates a plurality of logical devices, wherein the controller 12 has a shared memory 123 stored in an erase hold flag, and when an erase all command of an individual logical track of the logical device is received from the host device 10, the controller folds execution of the erase all, records this holding state in the erase hold flag, and when a format write request for setting a new data set is received later, the controller resets the erase holding flag and executes erase processing until the end of the logical track, so that control logic for preventing redundant and wasteful erase data writing is implemented.

【Claims】

【Claim 1】 A control method for an external storage sub-

system which comprises a plurality of rotary storage devices, and a controller having a cache memory for temporarily storing data exchanged between said rotary storage device and a host device, and emulates a plurality of logical devices for said host device, wherein when accesses of data from said host device to said logical devices are executed in a unit of a logical track which is separated into a plurality of unit data, and distributed and stored in a plurality of said rotary storage device, the erase operation, until the end of said logical track, is held when said host device instructs erasing for said logical track.

【Claim 2】 An external storage sub-system comprising:

a plurality of rotary storage devices; and

a controller having a cache memory for temporarily storing data exchanged between said rotary storage device and a host device,

wherein said controller emulates a plurality of logical devices for said host device, and executes accesses of data from said host device to said logical devices in a unit of a logical track which is separated into a plurality of unit data, and distributed and stored in a plurality of said rotary storage devices,

and wherein said controller further comprises a management flag for managing data in said individual logical track; and

a control logic for setting information that all the data on said logical track for said management flag

corresponding to said logical track is invalid, and holding the erase operation up to the end of the logical track when an erase instruction is instructed from the host device to an arbitrary logical track.

【Claim 3】 The external storage sub-system according to Claim 2, wherein said controller has control logic for performing at least one of the operations consisting of:

an operation to refer to said management flag when said host devices access one of said arbitrary logical tracks, and to process regarding no data in said logical track when said management flag is invalid;

an operation to clear said management flag and execute erase processing to the end of said logical track after format writing when said host device executes format writing for constructing a new data set to the logical track where said management flag is set, or when the power supply of said external storage sub-system is shut OFF;

an operation to record only said management flag on said rotary storage device without recording data corresponding to said logical track on said rotary storage device when data on said canceled logical track where said management flag is set and said management flag is removed from said cache memory, and to store the management information where said management flag is stored, to said cache memory and execute processing regarding no data for said access request from said host device when said host device requests access to said logical track; and

an operation to set said management flag to all of said logical tracks constituting said logical device and to hold the writing of predetermined format data in the initial format processing of said logical device which is executed before starting use from said host device, and to visualize and output the initialization status of data of said logical device when said initialization format is completed.

【Detailed Explanation of The Invention】

【0001】

【Field of The Invention】 The present invention relates to a control technology for an external storage sub-system and an external storage sub-system, and more particularly to a technology which is effective when applied to an external storage sub-system which stores the variable length data format received from a host device to a logical drive constructed on a rotary type storage device having a fixed length recording system redundant configuration.

【0002】

【Prior Art】 As stated in "Magnetic Disk Technique = Programming Handbook =", pp. 45 - 67, Version. 3, April 30, 1983, published by Takeuchi Book Store for example, when format writing is executed, a disk device for recording a variable length data format generally stores the data in a predetermined area, and also erases the data after recording to the end of the track to cancel the data.

【0003】 To erase the data of a logical device stored on an

external storage device via a host device, generally the following two means are taken. One is canceling the positional information of the data set recorded at the beginning of the logical device by format writing.

【0004】 With this method, however, data with an already canceled track may be referred to for a program which does not refer to the positional information of the data set. Therefore, another means is canceling all the data of the erase target data set in the logical device by format writing.

【0005】 For a command chain for canceling the user data of an entire track issued by the host device, it is reasonable to issue a format write command (Write R0) to the logical record number 0 which the user does not normally use, or to execute format write (End Of File Record) where the key length and the data length for the logical record number 1 are zero. Since this is a processing for canceling data of an entire track, a data length to minimize the data transfer time with the host device is preferable.

【0006】 Conventional erase processing after the end of format writing is often executed immediately after format writing, so if additional format writing is executed for this track, erase processing after format write processing is executed again, which is redundant and wasteful. A method to solve this technical problem is, for example, the technology reported on in Japanese Patent Laid-Open No. S61-241824. With the technology reported on in Japanese Patent

Laid-Open No. S61-241824, a table is set in a memory of the disk device for each logical track with respect to an erase operation after format writing ends, and at format writing, this table is dynamically referred to so as to omit unnecessary erase processing out of all erase processing when format writing ends, and to execute format writing efficiently.

【0007】

【Problems to be solved by the Invention】 The above mentioned prior art is very effective to omit unnecessary erase processing after format writing which is executed to add a new logical record with respect to the logical track where format has been written. However, in the prior art, when data is created again after format writing, to cancel the entire logical track, is executed, erase must be executed again. Also in the case of a disk array device which emulates a plurality of logical devices, it is physically impossible to store a control table on a disk device.

【0008】 Also when the cancellation of a logical volume is executed for a plurality of logical devices on a disk array device, the data transfer time with the host device is minimized, as mentioned above, but to record data on the disk device, the erased data is also included. Thereby, it becomes a major technical problem in terms of performance when the logical device has competition with a disk device in the array group.

【0009】 It is an object of the present invention to provide an external storage sub-system which can improve performance by optimizing erase processing without damaging the reliability of operation, and control technology thereof.

【0010】 It is another object of the present invention to provide an external storage sub-system which can improve the throughput of data set erasing in the storage device with a redundant storage configuration for emulating a plurality of logical devices, and control technology thereof.

【0011】 It is still another object of the present invention to provide an external storage sub-system which can decrease the time required for initialization formatting of a logical device, optimizing the management and configuration of the initialization status, and control technology thereof.

【0012】

【Means of Solving the Problems】 In the present invention, a control device has means for recognizing the command chain to be used for data set erasing, and a 1 bit management flag, for example, is set for each logical track. And when data set erasing is recognized, erase processing for this track is held until data set construction (format writing to the logical record No. 1, for example) is executed, where valid/invalid of data on the logical track is judged by the setting or cancellation status of this management flag, and access from the host device to the logical track is controlled.

【0013】 When data on the cancelled track is reflected from

the cache memory to the disk device, the present invention is used as means for guaranteeing that data on the track is invalid by recording only the management flag information without recording the data itself.

【0014】 When an initialization format is executed on the logical medium (device), only the management flag is set for each logical track, and actual data writing for formatting is held. If necessary, the management flag setting status in the initialization format is visualized and output to such a terminal as a service processor.

【0015】 In this way, according to the present invention, erasing on the track is held until a command chain for constructing a data set again is issued from the host device side, by considering the features of the command chain to be used for deleting data set, so processing time can be shortened by optimizing data set erasing. The data transfer processing volume, which is handled by the command chain used for data set erasing with the host device, is obviously smaller than that of the case when format writing is executed again. Therefore, data volume for erasing decreases when erasing is executed in a remaining area of a normal data write by format writing for constructing a data set to be subsequently processed, rather than executing erasing during data set erasing, which is clearly reasonable.

【0016】

【Preferred Embodiments】 Embodiments of the present invention will now be described with reference to the

accompanying drawings.

【0017】 Fig. 1 is a block diagram depicting an example of the configuration of an information processing system including an external storage sub-system which is an embodiment of the present invention, Fig. 2 is a conceptual diagram depicting an example of the management information to be used for an external storage sub-system of the present invention, Fig. 3 is a conceptual diagram depicting an example of the data storage method in the external storage sub-system of the present embodiment, and Fig. 4 is a conceptual diagram depicting an example of the visualization and output of the management information used for the external storage sub-system of the present embodiment. And Fig. 5 and Fig. 6 are flow charts depicting examples of the functions of the control method of the external storage sub-system of the present embodiment.

【0018】 The information processing system of the present embodiment in Fig. 1 comprises a host device 10, a channel 11 for controlling the input/output of data thereof, and external storage sub-systems.

【0019】 The external storage sub-system of the present embodiment is comprised of a storage device 13 constituted of a plurality of disk devices 13a, and a controller 12, which controls the data transfer between the storage devices 13 and the channel 11.

【0020】 In the case of the present embodiment, as shown in Fig. 3, the plurality of disk devices 13a of the storage

device 13 constitute a disk array, where a plurality of logical devices 13b are set. An individual logical device 13b includes a plurality of logical tracks 13c. And each one of the logical tracks 13c is comprised of a plurality of unit data 13d, which is distributed and stored in a plurality of disk devices 13a, and redundant data 13e, such as parity information, generated from this unit data 13d, to make up the redundant storage configuration of RAID 5. Because of this, when a failure occurs to one unit data 13d in the individual logical track 13c, the failed data can be restored from the other intact unit data 13d and redundant data 13e. In the case of the present embodiment, data is stored in a CKD (variable length recording) system, for example, for the logical track 13c of the individual logical device 13b.

【0021】 The controller 12 is comprised of a processor H121 which performs processing when an input/output request is received from the channel 11, a processor D122 which performs storage and reading control of data of the storage device 13, a data transfer block H124 for transferring data from the channel 11, a data transfer block D125 for transferring data from the storage device 13, a cache memory 126 where data transferred between the storage device 13 and the channel 11 is temporarily stored in a unit of the logical track 13c, for example, and a shared memory 123 which is accessed by a plurality of processors H121 and D122, and stores management information to execute exclusive

accompanying drawings.

【0017】 Fig. 1 is a block diagram depicting an example of the configuration of an information processing system including an external storage sub-system which is an embodiment of the present invention, Fig. 2 is a conceptual diagram depicting an example of the management information to be used for an external storage sub-system of the present invention, Fig. 3 is a conceptual diagram depicting an example of the data storage method in the external storage sub-system of the present embodiment, and Fig. 4 is a conceptual diagram depicting an example of the visualization and output of the management information used for the external storage sub-system of the present embodiment. And Fig. 5 and Fig. 6 are flow charts depicting examples of the functions of the control method of the external storage sub-system of the present embodiment.

【0018】 The information processing system of the present embodiment in Fig. 1 comprises a host device 10, a channel 11 for controlling the input/output of data thereof, and external storage sub-systems.

【0019】 The external storage sub-system of the present embodiment is comprised of a storage device 13 constituted of a plurality of disk devices 13a, and a controller 12, which controls the data transfer between the storage devices 13 and the channel 11.

【0020】 In the case of the present embodiment, as shown in Fig. 3, the plurality of disk devices 13a of the storage

device 13 constitute a disk array, where a plurality of logical devices 13b are set. An individual logical device 13b includes a plurality of logical tracks 13c. And each one of the logical tracks 13c is comprised of a plurality of unit data 13d, which is distributed and stored in a plurality of disk devices 13a, and redundant data 13e, such as parity information, generated from this unit data 13d, to make up the redundant storage configuration of RAID 5. Because of this, when a failure occurs to one unit data 13d in the individual logical track 13c, the failed data can be restored from the other intact unit data 13d and redundant data 13e. In the case of the present embodiment, data is stored in a CKD (variable length recording) system, for example, for the logical track 13c of the individual logical device 13b.

【0021】 The controller 12 is comprised of a processor H121 which performs processing when an input/output request is received from the channel 11, a processor D122 which performs storage and reading control of data of the storage device 13, a data transfer block H124 for transferring data from the channel 11, a data transfer block D125 for transferring data from the storage device 13, a cache memory 126 where data transferred between the storage device 13 and the channel 11 is temporarily stored in a unit of the logical track 13c, for example, and a shared memory 123 which is accessed by a plurality of processors H121 and D122, and stores management information to execute exclusive

control thereof.

【0022】 In the case of the present embodiment, a track management table 127, loaded from the storage device 13, is stored in the shared memory 123 when necessary.

【0023】 As Fig. 2 shows, the track management table 127 of the present embodiment stores track management information 127a on the logical track, such as track management information for controlling a disk array, and a 1 bit erase holding flag 127b for identifying whether an erase operation is held or not for the logical track.

【0024】 A service processor 14 having such a user interface as a display 14a, a keyboard 14b, and a mouse 14c is connected to the controller 12 according to need, so that a maintenance and management operator can monitor and control the operation of the controller 12 from the outside.

【0025】 An example of the functions of the present embodiment will now be described. When the disk array is comprised of a plurality of disk devices 13a of the storage device 13 so as to emulate a plurality of logical devices, initialization processing for an individual logical device 13b must be executed by the host device 10 before start of use.

【0026】 In this initialization processing, the initialization format of the logical track 13c of the emulated logical device 13b is executed, but in the case of the present embodiment, all that need be performed at this time is to set an erase holding flag 127b for all the

logical tracks of the logical device 13b, and the writing of initialization data is omitted. When an input/output request is received from the host device 10, it is possible to judge that data does not exist (already initialized) merely by referring to the erase holding flag 127b.

【0027】 If necessary, the maintenance and management operator of the system can visually check the initialization status (status of the erase holding flag 127b) of each logical device 13b by visualizing and outputting the status on the display 14a of the service processor 14, as shown in Fig. 4.

【0028】 The controller 12 shown in the present embodiment executes processing in the procedure shown in Fig. 5 when format writing to cancel an entire track is issued.

【0029】 When an input/output request is issued to a logical track, the controller 12 executes Step 300 for understanding the existence of data on this track. And the controller 12 judges whether the track management table 127 of the target logical track exists on the shared memory 123 in Step 310, and if no, Step 315 is executed to load the processing target track management table 127 to the shared memory 123. Then it is judged whether the command is a format writing command or not in Step 320.

【0030】 If the result of Step 320 is Yes, whether the access involves erasing the data of the logical track is judged in subsequent Steps 330, 340 and 350. In Step 330, whether the operation, specified by the Locate Record or the Locate

Record Extended command, is the Write Tracks operation, which executes erasing to the end of the track after 8 bytes of data block of the standard R0 record are transferred, is judged. In Step 340, whether the command is the Write R0 command for executing format writing to the R0 record is judged, and in Step 350, whether the command is format writing (Write CKD command or Erase command) for the logical record No. 1 and the data of the count block received from channel 11 has a 0 byte key length or a 0 byte data length, is judged. If the result is Yes in Steps 330, 340 and 350, the command can be recognized as a format writing command to cancel the data of an entire track, so processing to turn the erase holding flag 127b ON is executed in Step 360, and processing of Write related commands is executed in Step 370. At the time of Step 370, whether the format write command is chained or not is not clarified by the pattern of the command chain issued by the channel 11, except for the Write Track operation when the result of Step 330 is Yes, so only processing for a record to be processed by this command is executed, and erasing of data thereafter remains unexecuted.

【0031】 When the judge in Step 320 is No, that is, when the command is not a format writing command, and when the result of the judge executed in Step 330, Step 340 and Step 350 is No, even if the command is a format writing command, then judge in Step 400, shown in Fig. 6, is executed.

【0032】 Fig. 6 shows the flow of processing of commands that are not format writing commands for erasing data of an

entire track.

【0033】 In Step 400, whether the erase holding flag 127b is ON is judged for a track to be accessed or not. If No in Step 400, that is, if the erase holding flag 127b is not ON, then subsequent processing is omitted since it is sufficient to merely execute normal command processing.

【0034】 If the judge in Step 400 is Yes, then whether the command is a format writing command or not is judged in Step 410. If the command is not a format writing command, command processing is executed in Step 415 regarding that there is no data on this track. This is because some commands target the data of the next track. Although this is not shown in Fig. 6, it is clear that when an HA block or R0 record is the processing target, such as in the case of a Read HA command and Read Record Zero command, the command is accepted normally. Even when an R1 record is the processing target, processing can be continued regarding this as an End of File if key length and data length are 0 bytes.

【0035】 When the judgment in Step 410 is Yes, that is, when a format write command for constructing a new data set is issued, the erase holding flag 127b of this logical track is turned OFF in Step 411, and format write processing is executed in Step 412. In this format write processing in Step 412, processing to write a specified erase data until the end of the logical track is executed.

【0036】 If an erase has been executed for a logical track on the cache memory 126, only the erase holding flag 127b

related to this logical track is mirrored on the logical device 13b of the storage device 13 during the operation to mirror (transfer) the data to the logical device 13b on the storage device 13 of the logical track, where the operation to mirror data of the logical track itself is not executed. When an access request to the mirrored logical track is generated to the logical device 13b in this way from such a host device as a channel 11, the erase holding flag 127b on this logical track is read from the storage device 13, and is set in the track management table 127, and a no data status is responded for the erased logical track.

【0037】 According to the present embodiment, the erase holding flag 127b is set in the track management table 127, so in the format writing processing to cancel the data of an entire logical track, it is sufficient to merely execute the setting of the erase holding flag 127b and data transfer processing with the channel 11, therefore redundant erase data writing processing is prevented, the processing speed of data set erasing can be dramatically increased, throughput of data set erasing is improved, and the response of no data existing can be returned accurately to the access request from the host device 10 to a track which is recognized as erased by referring to the erase holding flag 127b, and as a consequence a reliability equivalent to a conventional control method can be secured.

【0038】 As a result, even when similar jobs, which are multiplexed, are executed, for example, a stable response

time can be obtained.

【0039】 Particularly when the erase of a logical track 13c is simultaneously executed on a plurality of logical devices 13b in a system where a plurality of logical devices 13b is set in the storage device 13 by RAID technology for example, access competition occurs at the physical level of the storage device in this status, and the time required for erasing dissipates depending on the logical address, but in the case of the present invention, erase processing can be executed at a predetermined response time regardless the logical address of a logical track 13c to be erased.

【0040】 The invention of the present inventors has been specifically described above based on the embodiment, but the present invention is not restricted by the above embodiment, and various changes can be made within the scope of the present invention which do not depart from the spirit thereof.

【0041】

【Effect of the Invention】 According to the control method of the external storage sub-system of the present invention, performance can be improved by optimizing erase processing without affecting the reliability of operations.

【0042】 Also according to the control method of the external storage sub-system of the present invention, throughput of data set erasing can be improved in storage devices with a redundant storage configuration which emulate a plurality of logical devices.

【0043】 Also according to the control method of the external storage sub-system of the present invention, time required for executing initialization format of a logical device can be shortened, and the management and confirmation of the initialization status can be optimized.

【0044】 According to the external storage sub-system of the present invention, performance can be improved by optimizing erase processing without affecting the reliability of operations.

【0045】 Also according to the external storage sub-system of the present invention, throughput of data set erasing can be improved in storage devices with a redundant storage configuration which emulate a plurality of logical devices.

【0046】 Also according to the external storage sub-system of the present invention, time required for executing the initialization format of a logical device can be shortened, and the management and confirmation of the initialization status can be optimized.

【Brief Description of the Drawings】

Fig. 1 is a block diagram depicting an example of the configuration of an information processing system, including an external storage sub-system according to an embodiment of the present invention;

Fig. 2 is a conceptual diagram depicting an example of the management information used for an external storage sub-system according to an embodiment of the present invention;

Fig. 3 is a conceptual diagram depicting an example of

a data storage method of an external storage sub-system according to an embodiment of the present invention;

Fig. 4 is a conceptual diagram depicting an example of a visualization output of the management information used for an external storage sub-system according to an embodiment of the present invention;

Fig. 5 is a flow chart depicting an example of the functions of a control method for an external storage sub-system according to an embodiment of the present invention; and

Fig. 6 is a flow chart depicting an example of the functions of a control method for an external storage sub-system according to an embodiment of the present invention.

【Explanation of Reference Numerals and Signs】

10 host device
11 channel
12 controller
13 storage device
13a disk device
13b logical device
13c logical track
13d unit data
13e redundant data
14 service processor
14a display
14b keyboard
14c mouse

121 processor H

122 processor D

123 shared memory

124 data transfer block H

125 data transfer block D

126 cache memory

127 track management table

127a track management information

127b erase holding flag

(19)日本国特許庁(JP)　(12)公開特許公報（A）　(11)特許出願公開番号

特開平１１－１４３６４６

(43)公開日　平成11年(1999)5月28日

(51)Int. Cl.[6]		識別記号	FI		
G06F	3/06	302	G06F	3/06	302 H
		540			540
G11B	19/02	501	G11B	19/02	501 B

審査請求　未請求　請求項の数３　ＯＬ　（全８頁）

(21)出願番号　特願平9-304716

(22)出願日　平成9年(1997)11月6日

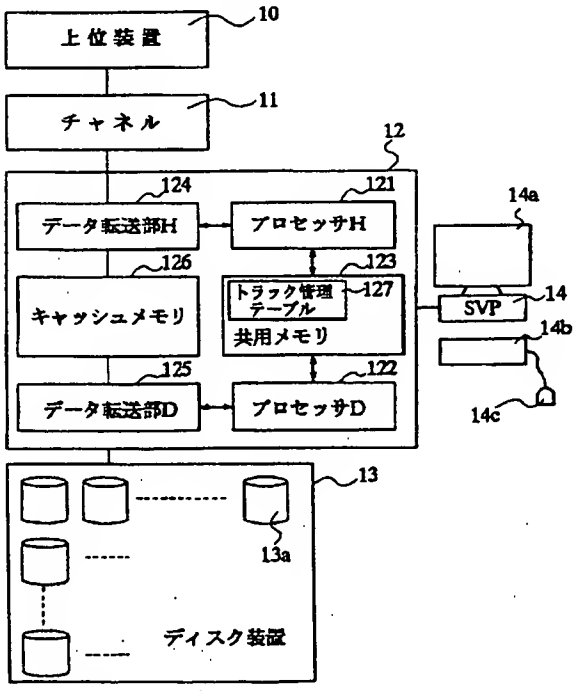
(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目6番地
(72)発明者　川口　勝洋
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内
(72)発明者　竹内　久治
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内
(72)発明者　黒川　勇
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内
(74)代理人　弁理士　筒井　大和

最終頁に続く

(54)【発明の名称】外部記憶サブシステムの制御方法および外部記憶サブシステム

(57)【要約】
【課題】　複数の論理デバイスをエミュレートする冗長記憶構成の記憶装置におけるデータセット消去のスループットを向上させる。
【解決手段】　上位装置１０と記憶装置１３との間に制御装置１２が配置され、記憶装置１３は、複数の論理デバイスをエミュレートする外部記憶サブシステムにおいて、制御装置１２の共用メモリ１２３内に、イレーズ保留フラグを設け、論理デバイス内の個々の論理トラックの全消去コマンドを上位装置１０から受領した時に、当該全消去の実行を保留するとともに、当該保留状態をイレーズ保留フラグに記録し、後の新規なデータセットの設定のための形式書き込み要求の受領時に、イレーズ保留フラグをリセットするとともに、論理トラック終端までのイレーズ処理を実行することで、重複した無駄なイレーズデータの書き込みを回避する制御論理を備えた。

図　1



1

【特許請求の範囲】

【請求項１】　複数の回転型記憶装置と、前記回転型記憶装置と上位装置との間で授受されるデータが一時的に格納されるキャッシュメモリを備えた制御装置とを含み、前記上位装置に対して複数の論理デバイスをエミュレートする外部記憶サブシステムの制御方法であって、前記上位装置から前記論理デバイスのデータへのアクセスは、各々が複数の単位データに分割されて複数の前記回転型記憶装置に分散して格納された論理トラック単位に実行されるとき、前記上位装置から前記論理トラックに対して消去指示があった場合、当該論理トラックの終端までの消去動作を保留することを特徴とする外部記憶サブシステムの制御方法。

【請求項２】　複数の回転型記憶装置と、前記回転型記憶装置と上位装置との間で授受されるデータが一時的に格納されるキャッシュメモリを備えた制御装置とを含み、前記制御装置は、前記上位装置に対して複数の論理デバイスをエミュレートし、前記上位装置から前記論理デバイスのデータへのアクセスは、各々が複数の単位データに分割されて複数の前記回転型記憶装置に分散して格納された論理トラック単位に実行されるようにした外部記憶サブシステムであって、

個々の前記論理トラック内のデータを管理するための管理フラグを備え、

前記制御装置は、前記上位装置から任意の前記論理トラックに対して消去指示があった場合、当該論理トラックに対応する前記管理フラグに対して、当該論理トラック内の全データが無効であることを示す情報を設定し、当該論理トラックの終端までの消去動作を保留する制御論理を備えたことを特徴とする外部記憶サブシステム。

【請求項３】　請求項２記載の外部記憶サブシステムにおいて、前記制御装置は、

前記上位装置から任意の前記論理トラックにアクセスがあった場合に、前記管理フラグを参照し、当該管理フラグが無効の場合には当該論理トラックにはデータ無しと見做して処理する操作、

前記管理フラグが設定されている論理トラックに対して、前記上位装置から新たなデータセットの構築のための形式書き込みが実施された場合、または前記外部記憶サブシステムの電源切断時に、前記管理フラグを解除し、当該形式書込み処理後に当該論理トラックの終端までのイレーズ処理を実行する操作、

前記管理フラグが設定されている無効化された前記論理トラックのデータおよび前記管理フラグを前記キャッシュメモリから追い出す場合に、当該論理トラックに対応するデータは前記回転型記憶装置に記録せず、前記管理フラグのみを前記回転型記憶装置上に記録し、前記上位装置から当該論理トラックにアクセス要求があった場合には、前記管理フラグが記憶されている管理情報を前記キャッシュメモリに格納し、前記上位装置からの前記アクセス要求に対して、データ無しとして処理を実行する操作、

前記上位装置から使用を開始する前に実施される前記論理デバイスの初期フォーマット処理において、当該論理デバイスを構成する全ての前記論理トラックに対して前記管理フラグを設定するとともに特定のフォーマットデータの書き込みは保留し、前記初期フォーマットが完了した場合に、当該論理デバイスのデータが初期化されている状態を外部に可視化して出力する操作、

の少なくとも一つの操作を行う制御論理を備えたことを特徴とする外部記憶サブシステム。

2

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、外部記憶サブシステムの制御技術および外部記憶サブシステムに関し、特に、たとえば上位装置から受領した可変長のデータ形式を、固定長記録方式の冗長構成の回転型記憶装置上に構築された論理ドライブに格納する外部記憶サブシステム等に適用して有効な技術に関する。

【０００２】

【従来の技術】たとえば、株式会社竹内書店新社、１９８３年４月３０日第３刷発行「磁気ディスク技法集＝プログラミング・ハンドブック＝」Ｐ４５～Ｐ６７、等の文献にも記載されているように、一般に可変長のデータ形式を記録するディスク装置は、形式書込みが実行された場合に、当該データを所定のエリアに格納すると共に当該レコード以降のデータをトラック終端までイレーズを行い、データを無効化する。

【０００３】上位装置を介在させて、外部記憶装置に格納された論理デバイスのデータを消去する場合には、以下の２つの手段が一般的である。一つは、当該論理デバイスの先頭に記録されているデータセットの位置情報を形式書込みにより無効化することである。

【０００４】しかし、この方法では、データセットの位置情報を参照しないプログラムでは、既に無効化されたトラックのデータを参照する可能性がある。このため、もう一つの手段として、当該論理デバイスで消去対象のデータセット全てのデータを形式書込みにより無効化することが考えられる。

【０００５】上位装置が発行するトラック全体のユーザデータを無効化するコマンドチェインは、通常はユーザが使用しない論理レコード番号０番に対して形式書込みのコマンド（Ｗｒｉｔｅ　Ｒ０）を発行するか、論理レコード番号１番に対して、キー長およびデータ長がゼロの形式書込み（Ｅｎｄ　Ｏｆ　Ｆｉｌｅレコード）を実行することが合理的である。これは、全トラックのデータを無効化するための処理であるため、上位装置とのデータ転送時間を最短にするデータ長が望ましいためである。

【０００６】従来の形式書込み終了後のイレーズ処理

は、形式書込み直後に行われることが多いため、当該トラックに対して更に形式書込みが行われた場合、再度形式書込み処理後のイレーズ処理が実行されることになり、重複して無駄なイレーズを実施することになる。この技術的課題を解決するための方法として、例えば、特開昭６１－２４１８２４号の技術が挙げられる。この特開昭６１－２４１８２４号の技術では、形式書込み終了後のイレーズ動作に関して各論理トラック毎にテーブルをディスク装置内のメモリに設け、形式書込み時にダイナミックにこのテーブルを参照して、形式書込み終了時のイレーズ処理のうち、不要なイレーズ処理を省くことにより、効率良く形式書込みを実施しようとするものである。

【０００７】

【発明が解決しようとする課題】上記の従来技術は形式書込みされた当該論理トラックに対して、新たな論理レコードを追加するために実施する形式書込み後の不要なイレーズ処理を省くことは非常に有効であるが、当該論理トラック全体を無効化する形式書込み後、再度データセットが再作成されるような場合には再度イレーズが実行されることになる。また、複数の論理デバイスをエミュレートしているディスクアレイ装置においては、制御テーブルをディスク装置内に記憶しておくことは、物理的に不可能である。

【０００８】また、ディスクアレイ装置において、複数の論理デバイスに対して論理ボリュームの無効化が実行された場合、上位装置とのデータ転送時間は前述のように最小化されるが、ディスク装置にデータを記録しようとした場合には、通常イレーズされたデータも含まれるため、論理デバイスがアレイグループ内のディスク装置と競合するような場合には性能上大きな技術的課題となることが懸念される。

【０００９】本発明の目的は、動作の信頼性を損なうことなく、イレーズ処理の最適化による性能向上を実現することが可能な外部記憶サブシステムおよびその制御技術を提供することにある。

【００１０】本発明の他の目的は、複数の論理デバイスをエミュレートする冗長記憶構成の記憶装置におけるデータセット消去のスループットを向上させることが可能な外部記憶サブシステムおよびその制御技術を提供することにある。

【００１１】本発明の他の目的は、論理デバイスの初期化フォーマットにおける所要時間の短縮や初期化状態の管理および確認の的確化を実現することが可能な外部記憶サブシステムおよびその制御技術を提供することにある。

【００１２】

【課題を解決するための手段】本発明では、データセット消去に使用されるであろうコマンドチェインを制御装置が認識する手段を設けるとともに、個々の論理トラック毎にたとえば１ビットの管理フラグを設定する。そして、データセット消去と認識した場合には、再度、データセットの構築（に伴う、たとえば論理レコード番号１番への形式書き込み）が実行されるまでは、当該トラックに対するイレーズ処理を保留し、本管理フラグの設定、解除状態により当該論理トラックのデータの有効／無効を判定して上位装置からの当該論理トラックへのアクセスを制御する。

【００１３】また、無効化されたトラックのデータをキャッシュメモリからディスク装置に反映する場合に、データそのものは記録せず、当該管理フラグ情報のみを記録して、当該トラック内のデータが無効であることを保証するための手段として使用する。

【００１４】また、論理媒体（デバイス）の初期化フォーマット時には、個々の論理トラック毎に管理フラグをセットするだけで、フォーマット用の実際のデータ書き込みは保留する。また、必要に応じて、当該初期化フォーマットにおける管理フラグの設定状態を、サービスプロセッサ等の端末に可視化して出力する。

【００１５】このように、本発明では、通常データセットの消去に使用されるであろうコマンドチェインの特徴を考慮して、かかる場合には上位装置側から再度データセット構築のためのコマンドチェインが発行されるまでは当該トラックに対するイレーズを保留するので、データセット消去の最適化による処理時間の短縮を実現できる。データセット消去に使用されるコマンドチェインで扱う上位装置との間でのデータ転送処理量は、再度形式書込みが実施された場合のそれと比べて明らかに少ないため、データセット消去でイレーズを実施するよりも、後続で処理されるデータセットの構築のための形式書込みで通常のデータの書き込みの残りの領域でイレーズを実施した方が、イレーズ用のデータ量が減り、明らかに合理的である。

【００１６】

【発明の実施の形態】以下、本発明の実施の形態を図面を参照しながら詳細に説明する。

【００１７】図１は、本発明の一実施の形態である外部記憶サブシステムを含む情報処理システムの構成の一例を示すブロック図であり、図２は、本実施の形態の外部記憶サブシステムにて用いられる管理情報の一例を示す概念図、図３は、本実施の形態の外部記憶サブシステムにおけるデータ格納方法の一例を示す概念図、図４は、本実施の形態の外部記憶サブシステムにて用いられる管理情報の可視化出力の一例を示す概念図である。また、図５および図６は本実施の形態の外部記憶サブシステムの制御方法の作用の一例を示すフロチャートである。

【００１８】図１に示した本実施の形態の情報処理システムは、上位装置１０と、そのデータ入出制御を司るチャネル１１と、配下の外部記憶サブシステムからなる。

【００１９】本実施の形態の外部記憶サブシステムは、

複数のディスク装置１３ａから構成される記憶装置１３と、この記憶装置１３とチャネル１１との間に介在してデータ転送制御を行う制御装置１２から構成される。

【００２０】本実施の形態の場合、図３に例示されるように、一例として、記憶装置１３の複数のディスク装置１３ａはディスクアレイを構成し、複数の論理デバイス１３ｂが設定されている。個々の論理デバイス１３ｂは、複数の論理トラック１３ｃを含んでいる。さらに、個々の論理トラック１３ｃは、複数のディスク装置１３ａに分散して格納される複数の単位データ１３ｄおよびこれらの単位データ１３ｄから生成されたパリティ等の冗長データ１３ｅで構成され、たとえばＲＡＩＤ５の冗長記憶構成を採っている。これにより、個々の論理トラック１３ｃにおいて一つの単位データ１３ｄに障害が発生した時には、他の健全な単位データ１３ｄおよび冗長データ１３ｅから障害データの復元が可能となっている。また、本実施の形態の場合、個々の論理デバイス１３ｂの論理トラック１３ｃに対しては、一例としてＣＫＤ（可変長記録）方式にてデータが格納される。

【００２１】制御装置１２は主にチャネル１１からの入出力要求のあった場合の処理を行うプロセッサＨ１２１と、配下の記憶装置１３に対するデータの格納および読み出し制御を行うプロセッサＤ１２２と、チャネル１１からのデータ転送を行うデータ転送部Ｈ１２４と、記憶装置１３からのデータ転送を行うデータ転送部Ｄ１２５と、記憶装置１３とチャネル１１との間で授受されるデータ等が、たとえば論理トラック１３ｃの単位にて、一時的に格納されるキャッシュメモリ１２６と、複数のプロセッサＨ１２１およびプロセッサＤ１２２からアクセスされ、両者間での排他制御を行うための管理情報等が格納される共用メモリ１２３とから構成されている。

【００２２】本実施の形態の場合、共用メモリ１２３内には必要に応じて記憶装置１３から吸い上げたトラック管理テーブル１２７が格納される。

【００２３】図２に例示されるように、本実施の形態のトラック管理テーブル１２７には、個々の論理トラック毎に、当該論理トラックに関する、たとえばディスクアレイ制御等のためのトラック管理情報１２７ａの他に、当該論理トラックに対するイレーズ操作の保留の有無等を識別するための１ビットのイレーズ保留フラグ１２７ｂが格納されている。

【００２４】制御装置１２には、必要に応じて、ディスプレイ１４ａやキーボード１４ｂ、マウス１４ｃ、等のユーザインタフェースを備えたサービスプロセッサ１４が接続されており、外部から保守管理者等が、制御装置１２の動作の監視や制御を行うことが可能になっている。

【００２５】以下、本実施の形態の作用の一例を説明する。まず、本実施の形態のように、記憶装置１３の複数のディスク装置１３ａにてディスクアレイを構成し、複数の論理デバイスをエミュレートする場合には、上位装置１０により使用開始する前に、個々の論理デバイス１３ｂのイニシャライズ処理を実行する必要がある。

【００２６】このイニシャライズ処理では、エミュレートされた論理デバイス１３ｂの論理トラック１３ｃの初期化フォーマットを実施するが、この時に、本実施の形態の場合には、当該論理デバイス１３ｂの全論理トラックにイレーズ保留フラグ１２７ｂを設定するだけで、初期化データの書き込みは省略される。上位装置１０から入出力要求があった場合には、イレーズ保留フラグ１２７ｂを参照することでデータが存在しない（初期化済である）ことを一度で判断することが可能である。

【００２７】また、必要に応じて、図４に例示されるように、サービスプロセッサ１４のディスプレイ１４ａに、各論理デバイス１３ｂの初期化状態（イレーズ保留フラグ１２７ｂの状態）を可視化して出力することにより、システムの保守管理者が目視にて、確認することが可能になる。

【００２８】本実施の形態に示した制御装置１２は、図５に例示される手順でトラック全体を無効化する形式書込みが発行された場合の処理を行う。

【００２９】制御装置１２はある論理トラックに入出力要求が発行された場合に当該トラックのデータの存在を把握するためのステップ３００を実行する。そして、対象論理トラックのトラック管理テーブル１２７が共用メモリ１２３上に存在するか否かの判断をステップ３１０で行い、Ｎの場合には、処理対象のトラック管理テーブル１２７を共用メモリ１２３に吸い上げを行うステップ３１５を実行する。その後、形式書込みコマンドか否かの判断をステップ３２０で実行する。

【００３０】ステップ３２０でＹの場合には、次に論理トラックのデータをイレーズする様なアクセスか否かの判断を後続するステップ３３０、ステップ３４０およびステップ３５０で実施する。ステップ３３０ではＬｏｃａｔｅ　ＲｅｃｏｒｄまたはＬｏｃａｔｅ　Ｒｅｃｏｒｄ　Ｅｘｔｅｎｄｅｄコマンドで指定されるオペレーションが標準Ｒ０レコードのデータ部８バイトの転送後、当該データ転送以降はトラック終端までイレーズを実行するＷｒｉｔｅ　Ｔｒａｃｋｓオペレーションか否かの判断を行い、ステップ３４０ではＲ０レコードに対する形式書込みを行うＷｒｉｔｅ　Ｒ０コマンドであるか否かの判断を行い、ステップ３５０では論理レコード番号１番に対する形式書込み（Ｗｒｉｔｅ　ＣＫＤコマンドまたはＥｒａｓｅコマンド）でかつチャネル１１より受領するカウント部のデータがキー長が０バイト、データ長が０バイトであるかを判断する。これらステップ３３０、ステップ３４０およびステップ３５０でＹとなる場合にはトラック全体のデータを無効とする形式書込みコマンドとして認識することが可能であるため、イレーズ保留フラグ１２７ｂをＯＮにする処理をステップ３６０

で実施し、Ｗｒｉｔｅ系コマンドの処理をステップ３７０で行う。ステップ３７０の時点では形式書込みコマンドがチェインするか否かはステップ３３０がＹのＷｒｉｔｅ Ｔｒａｃｋオペレーションを除いてはチャネル１１より発行されるコマンドチェインのパタンによっては明示されないため、当該コマンドで処理するレコードの処理のみを実行し、それ以降のデータのイレーズは未実施にしておく。

【００３１】ステップ３２０の判定がＮ、すなわち形式書込みコマンドでない場合と、形式書込みコマンドであってもステップ３３０、ステップ３４０およびステップ３５０で実行される判断の結果がＮの場合には図６で示されるステップ４００の判定を行う。

【００３２】図６では、トラック全体のデータを消去するための形式書込みコマンド以外のコマンド処理の流れを示している。

【００３３】ステップ４００は、アクセスするトラックに対してイレーズ保留フラグ１２７ｂがＯＮになっているかを判断する。ステップ４００でＮ、すなわちイレーズ保留フラグ１２７ｂがＯＮで無い場合には、通常のコマンド処理を実行すればよいので、以降の処理については省略する。

【００３４】ステップ４００がＹの場合には、ステップ４１０で形式書込みコマンドであるか否かを判断する。形式書込み以外のコマンドならばステップ４１５で当該トラックのデータは無しとしてコマンド処理を行う。これは、コマンドの種別によっては、次トラックのデータを対象にするものが存在するためである。また、図６に示していないが、Ｒｅａｄ ＨＡコマンドやＲｅａｄ Ｒｅｃｏｒｄ ＺｅｒｏコマンドのようにＨＡ部やＲ０レコードを処理対象とする場合には正常にコマンドを受付けることは本発明の趣旨からも明白である。また、Ｒ１レコードが処理対象の場合でもキー長とデータ長が０バイトの場合にはＥｎｄ Ｏｆ Ｆｉｌｅとして処理を継続することも同様である。

【００３５】ステップ４１０がＹの場合、すなわち、新たなデータセットを構築するための形式書込みコマンドが発行された場合には、当該論理トラックのイレーズ保留フラグ１２７ｂをステップ４１１でＯＦＦし、ステップ４１２では形式書込み処理を実行する。なお、このステップ４１２の形式書込み処理では、論理トラックの末端まで特定のイレーズデータを書き込む処理が実行される。

【００３６】また、キャッシュメモリ１２６上の論理トラックに対してイレーズ実行されていた場合、当該論理トラックの記憶装置１３上の論理デバイス１３ｂに対して反映させる（追い出す）操作では、当該論理トラックに関するイレーズ保留フラグ１２７ｂのみを記憶装置１３上の論理デバイス１３ｂに反映させ、当該論理トラックのデータ自体を反映させる操作は行わない。また、こうして論理デバイス１３ｂに対して反映された当該論理トラックに対してチャネル１１等の上位装置からアクセス要求が発生した場合には、当該論理トラックに関するイレーズ保留フラグ１２７ｂを記憶装置１３から読出してトラック管理テーブル１２７に設定し、イレーズ済の論理トラックとして、データ無しを応答するような処理が行われる。

【００３７】このように本実施の形態によれば、トラック管理テーブル１２７内にイレーズ保留フラグ１２７ｂを設けることにより、論理トラック全体のデータを無効化する形式書込み処理では、イレーズ保留フラグ１２７ｂの設定とチャネル１１と行われるデータ転送処理のみ行えばよいので、重複したイレーズデータの書き込み処理が回避され、データセット消去の処理が大幅に高速化され、データセット消去のスループットが向上すると同時に、イレーズ保留フラグ１２７ｂの参照によって、イレーズ済みと認識されるトラックに対する上位装置１０からのアクセス要求に対して的確にデータ無しと応答することができるため、従来の制御方法と同等の信頼性を確保することが出来る。

【００３８】この結果、たとえば、同様のジョブが多重で実行された場合にも、安定したレスポンスタイムを得ることが可能になる。

【００３９】特に、たとえば、記憶装置１３にＲＡＩＤ技術等によって複数の論理デバイス１３ｂを設定して使用するシステムにおいて、複数の論理デバイス１３ｂにて同時に論理トラック１３ｃのイレーズを実行する場合、そのままでは物理的な記憶装置のレベルでのアクセス競合が発生して、論理アドレスによってイレーズ処理の所要時間にばらつきを生じるが、本実施の形態の場合には、イレーズ対象の論理トラック１３ｃの論理アドレスに関係なく、一定のレスポンスタイムにてイレーズ処理を行うことができる。

【００４０】以上本発明者によってなされた発明を実施の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【００４１】

【発明の効果】本発明の外部記憶サブシステムの制御方法によれば、動作の信頼性を損なうことなく、イレーズ処理の最適化による性能向上を実現することができる、という効果が得られる。

【００４２】また、本発明の外部記憶サブシステムの制御方法によれば、複数の論理デバイスをエミュレートする冗長記憶構成の記憶装置におけるデータセット消去のスループットを向上させることができる、という効果が得られる。

【００４３】また、本発明の外部記憶サブシステムの制御方法によれば、論理デバイスの初期化フォーマットにおける所要時間の短縮や初期化状態の管理および確認の

的確化を実現することができる、という効果が得られる。

【００４４】本発明の外部記憶サブシステムによれば、動作の信頼性を損なうことなく、イレーズ処理の最適化による性能向上を実現することができる、という効果が得られる。

【００４５】また、本発明の外部記憶サブシステムによれば、複数の論理デバイスをエミュレートする冗長記憶構成の記憶装置におけるデータセット消去のスループットを向上させることができる、という効果が得られる。

【００４６】また、本発明の外部記憶サブシステムによれば、論理デバイスの初期化フォーマットにおける所要時間の短縮や初期化状態の管理および確認の的確化を実現することができる、という効果が得られる。

【図面の簡単な説明】

【図１】本発明の一実施の形態である外部記憶サブシステムを含む情報処理システムの構成の一例を示すブロック図である。

【図２】本発明の一実施の形態である外部記憶サブシステムにて用いられる管理情報の一例を示す概念図である。

【図３】本発明の一実施の形態である外部記憶サブシステムにおけるデータ格納方法の一例を示す概念図である。

【図４】本発明の一実施の形態である外部記憶サブシステムにて用いられる管理情報の可視化出力の一例を示す概念図である。

【図５】本発明の一実施の形態である外部記憶サブシステムの制御方法の作用の一例を示すフロチャートである。

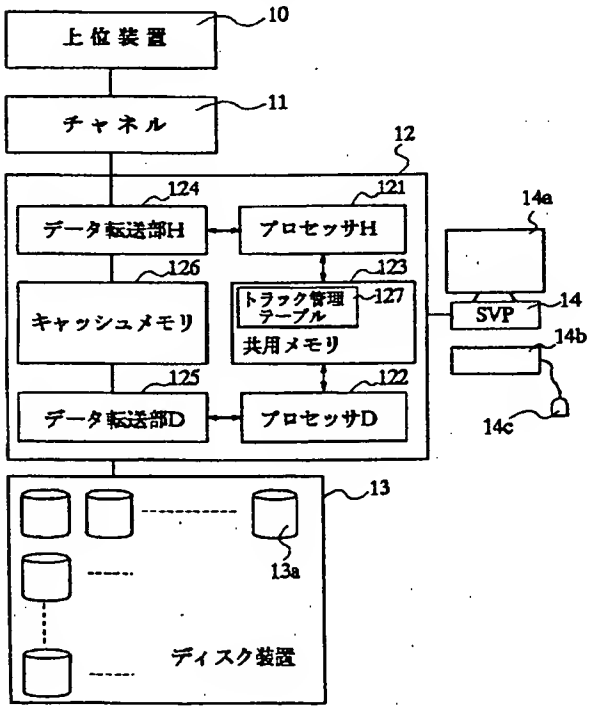
【図６】本発明の一実施の形態である外部記憶サブシステムの制御方法の作用の一例を示すフロチャートである。

【符号の説明】

１０…上位装置、１１…チャネル、１２…制御装置、１３…記憶装置、１３ａ…ディスク装置、１３ｂ…論理デバイス、１３ｃ…論理トラック、１３ｄ…単位データ、１３ｅ…冗長データ、１４…サービスプロセッサ、１４ａ…ディスプレイ、１４ｂ…キーボード、１４ｃ…マウス、１２１…プロセッサH、１２２…プロセッサD、１２３…共用メモリ、１２４…データ転送部H、１２５…データ転送部D、１２６…キャッシュメモリ、１２７…トラック管理テーブル、１２７ａ…トラック管理情報、１２７ｂ…イレーズ保留フラグ。

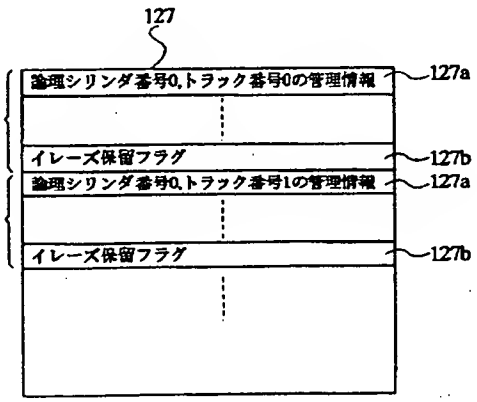
【図１】

図　1



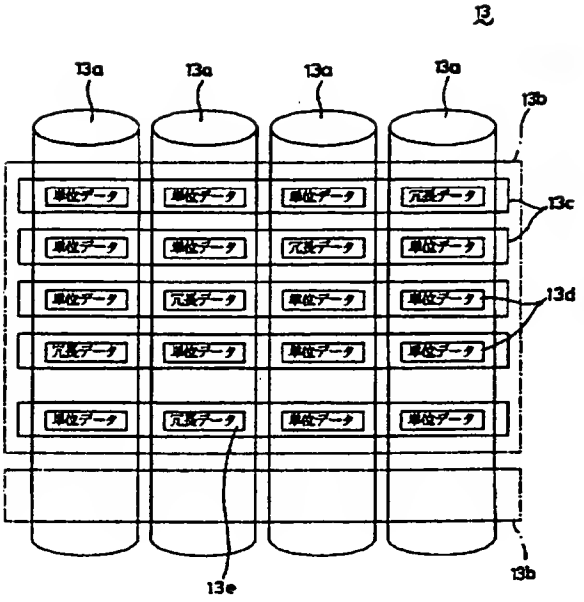
【図２】

図　2



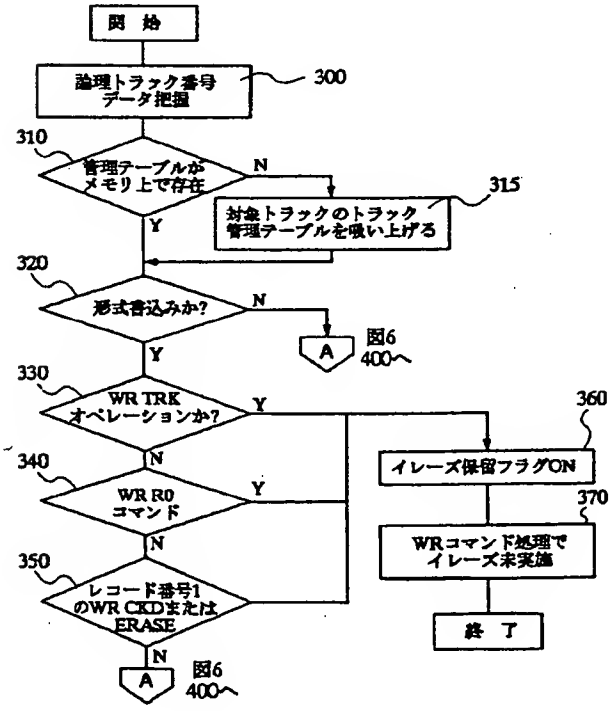
【図３】

図３



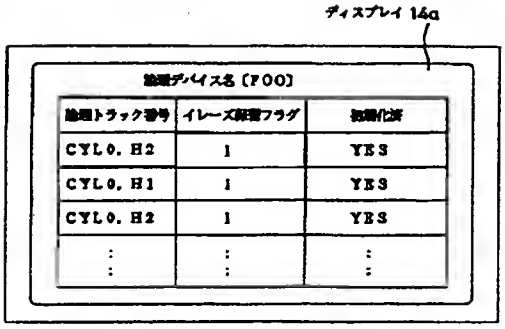
【図５】

図　５



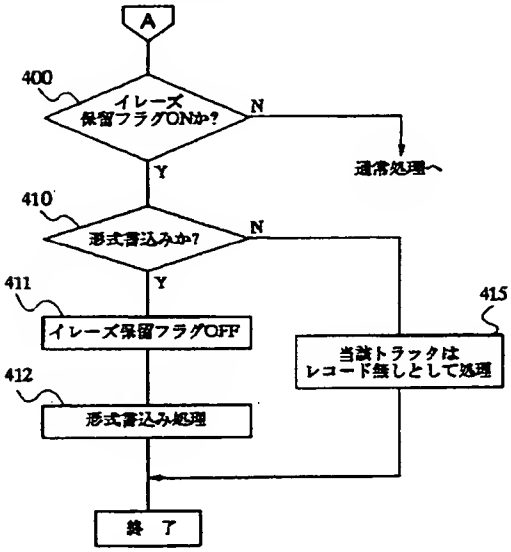
【図４】

図４



【図６】

図　６



フロントページの続き

(72)発明者　小沼　弘明
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内

(72)発明者　和田　賢一
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内